NIH Center for Scientific Review

peer review notes

september 2013

In this Issue

Reviewing Discovery-Based Research

Discovery- or non-hypothesis-based research has tremendous potential to advance fields of research where little is known or where the development of innovative tools and technologies and the gathering of new information could help open up new areas of research and allow researchers to test new hypotheses.

Most NIH research projects are designed to test hypotheses, so many scientists think it is challenging to apply for and review discovery-based research. We thus decided to discuss this important type of research and provide some guidance to the community.

The Explosion of Discovery-Based Research

With dramatic advances in biotechnology, including bioengineering and nanotechnology, NIH has received a rapidly growing number of non-hypothesis-based research applications. Technology development may take many forms, including the development of devices, imaging technologies, novel uses for existing technologies as well as methods for data storage, retrieval and dissemination.

Discovery-based research also can take the form of screening projects for drug discovery as well as genetic research projects including the high-profile Human Genome Project and Genome-Wide Association Studies (GWAS). The recently announced Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative also will support discovery-based research.

Large population scientists also may propose high-value discovery-based research, such as assembling a unique cohort—e.g., one for an important but previously unstudied phenotype—or a set of epidemiological data that could enable others to advance many promising hypothesis-driven research projects.

What Reviewers Should Consider

Reviewing discovery-based research may require reviewers to shift gears. To help you meet this challenge, we offer the following suggestions:

- ***Remember the main goals for NIH review and research*.** Study sections are charged with assessing the likelihood that the project will exert a sustained and powerful influence. For discovery research, the proposed research may influence multiple fields in specific ways that may not be readily anticipated, but you may be able to assess the general potentials.

- ***Be prepared to adjust your mindset of focusing only on near-term answers for immediate questions.*** You need to remember that discovery research applications require that you appraise the plausible scientific impact of such exploratory research projects with a 5-10 year time horizon. Your enthusiasm should thus be driven by your consideration of an application's impact on the field. Whether or not an application is hypothesis- or discovery-driven should not drive your enthusiasm.

- ***Stay grounded:*** Your strong foundation in hypothesis-driven research will serve you well as you discuss how the proposed research may expand the knowledge and technology bases for future research or generate new tools or data that could spawn new hypotheses and scientific insights.

Specific Things a Reviewer *Might* Want to Consider

- ***Did the application note the importance of the biological or behavioral context?*** While research designed to develop new technologies or resources does not have to be hypothesis-driven in the traditional sense, a clear biological or behavioral context is important, such as a new imaging technique to "see" a known biologic change that could lead to further studies and advances in the field.

- ***Has the application included biomedical experts early in the development of technology anticipated to be useful in a medical setting?*** While many uses for a given technology can be envisioned, plans for testing in a specific research area can help to focus the research effort. Including appropriate biological expertise could help guide development of the technology as it progresses.

- ***Did the application provide plans for how data will be analyzed and selected for further study?*** Such plans may be useful if the discovery-based

research seeks to generate data related to complex disorders or signaling pathways.

- ***Did the application explain how the research may address current barriers?*** And how could expected outcomes advance scientific knowledge or clinical practice in a specific area of research?

- ***Did the application discuss strategies for managing risk as the project proceeds?*** Non-hypothesis driven research may be seen as riskier than traditional research projects, so doing this may be important, particularly for research with a long project period.

How Not to Miss Important Information about Receipt and Referral of Your Application

CSR's Division of Receipt and Referral processes over 80,000 NIH grant applications each year. Sometimes during this process, we may need to give you critical information about the status of your application. We might need to tell you we cannot honor your request for a study section or institute assignment. Or we might need to communicate time-sensitive action items, such as requests for additional material needed to move your application forward in the review process.

To help you navigate such bumps in the road to review, we collaborated with eRA Commons to produce a new video that will help you see these important notifications.

Key Points You Need to Know

- Email notifications will be sent to you from era-notify@mail.nih.gov. Make sure "era-notify" is in your trusted senders list so that these important emails are not blocked by your spam filter.

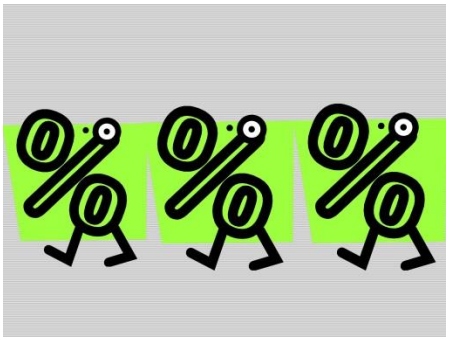
- The email address we use is the last one you entered in your eRA Commons account profile. Be sure your account is updated with your current email address.

- Email is not 100% reliable but your status screen in eRA Commons is always updated with the latest information. Please check the status screen for each application in eRA Commons periodically throughout the receipt, referral and review process to ensure that you receive all important notifications.

View the Video

Watch the new NIH video tutorial on How Not to Miss Important Information about Receipt and Referral of Your Application. The video uses screenshots and step-by-step instructions to guide you from your landing page in eRA Commons right to the information you need to know.

Understanding Percentiles of Scored Grant Applications



The different study sections that review NIH grant applications tend to adopt different scoring perspectives. Some study sections have a generous "Lake Wobegon" approach—most of their applications score above the NIH average. Other study sections may take a more "Tough Love" approach—very few of their applications receive exceptional impact scores.

If NIH funded according to absolute impact scores, the Lake Wobegon applications would have an unfair advantage. NIH has a longstanding practice of percentiling applications to normalize scoring across study sections so funding is more evenly distributed.

How Do We Generate Percentiles for R01s?

- ***Study section reviewers assign an overall impact score t***o each application they discuss during the meeting, scoring in whole numbers from 1 to 9.

- ***We average the impact scores assigned to each application,*** multiply by 10, and round to the nearest integer. For example, a 1.34 average becomes 13.

- ***We enter the final impact scores for all of the R01 applications a regular study section has reviewed into a rank-ordered percentiling table, along with all of the R01 scores from the previous two review rounds.*** Combining scores in this way reduces the effects of minor variations over time. Not Discussed applications are included with an impact score equivalent to 91, putting them at the bottom of the table.

- ***We calculate percentiles using the following formula:*** Percentile= 100 x (rank# minus 0.5) divided by the total number of applications in the table.

- ***We then round up the percentile score to a whole number*** (10.1 becomes 11) so percentiles range from 1 to 100. The percentile tells us the percentage of the applications that received a score equal to or better than that specific application.

Percentiling Artifacts

What happens when reviews don't follow the "norm"? When we percentile, we assume high quality applications are fairly equally distributed across scientific review groups. But this may not always be the case. Percentiling creates equal steps between applications even if there are big gaps in absolute scores in a given review meeting. For instance, a study section may give exceptional scores to three applications and find that the three next best applications were only above average. When these latter applications are percentiled, they might receive better percentile scores than the reviewers would have intended.

Fortunately, this isn't a big problem because our study sections tend to review large numbers of applications at a time, so the percentile steps, and these distortions between applications, are relatively small. To prevent significant distortions, CSR monitors the numbers of applications its study sections review and works to reorganize them when the numbers of applications they review fall significantly.

What Happens When Applications Receive Scores that Tie? Reviewers will often assign the same score to multiple applications. Although a rank-ordering process could give each of the tied applications a different percentile value, it is unfair to give different percentile values to applications in the same study section that received the same impact score. Instead, all of the *ranks* for tied applications are averaged, and used to give all of the tied applications the same percentile value. Note that this means that the percentile table takes a larger step when it gets to a group of tied scores, and when it leaves that group.

This is less of a problem when study sections review a large number of applications, as they typically do. Nonetheless, we encourage reviewers to spread their scores to reduce ties and provide more discrimination in their scorings so the NIH institutes and centers have more information to work with when they make funding decisions.

Percentiling Other Applications

R21 and R03 Applications: Summary statements for these applications often include a "shadow percentile" value, which is derived from the R01 percentiling table. Scores from these types of applications are given a percentile based on the percentile that would have been given to an R01 application with the same impact score. Since not all possible impact score values are present in the percentile table, percentiles for impact scores missing in the table are derived by interpolation within the table. It is important to note that non-R01 scores do not contribute to the percentile table, even if they are assigned "percentiles" from that table.

Applications Reviewed in CSR Special Emphasis Panels: When R01 or other applications that require percentiling are reviewed in nonrecurring special emphasis panels, we calculate percentiles using the percentile table based on all R01 applications reviewed in regular CSR study sections for the past three rounds.

Fellowship Applications are percentiled against a percentile table generated by the impact scores within each separate Fellowship study section, again combining three review rounds.

Many mechanisms, including SBIRs and STTRs, are not percentiled.

Fair and Useful Assessments

Percentiling certainly does a lot to help NIH make better decisions about which applications to fund by removing differences in study section scoring tendencies from this decision process. But like any system it isn't perfect.

To Counter Imperfections:

- CSR is working hard to encourage reviewers to use a broader range of scores for the applications in their meetings so reviewers can give clearer advice to NIH about which projects would represent the best investment of our limited research dollars.

- NIH is looking at ways to test the assumption that quality applications are evenly distributed among study sections.

This article expands on a post published on the NIH "Rock Talk" blog, where there is an ongoing discussion on percentiling.

CSR Hosts Seminar to Explore Ranking Grant Applications

CSR invited experts in voting and ranking to help us consider the potentials and pitfalls as we explore the usefulness of direct ranking of applications in peer-review. We currently use a system of absolute scores that are later converted to percentiles. The speakers brought a wide range of professional experience and theoretical acumen related to voting methods and how they might be applied to peer review. They brought to the table broad experiences such as civil juror decision-making, political elections, and the Grammy award process.

Speakers included Dr. Andrea Hollingshead of the University of Southern California; Dr. Reid Hastie of the University of Chicago; Dr. David Budescu of Fordham University; and Dr. Donald Saari of the University of California, Irvine.

The discussions were lively as the speakers considered the different methods of ranking as well as the potential pitfalls applying them to NIH peer-reviews. Participants also discussed different weighting schemes for criterion scores.

Challenges and Opportunities

Dr. Hollingshead opened the symposium with a discussion of the current NIH review system and how direct ranking might be added to the process to provide additional information to program staff for making funding decisions. She also discussed the consequences of different choices of rank order procedures. While Dr. Hollingshead noted that information exchange provided by peer review meetings adds value to the process, she questioned whether it is reasonable for the scores of assigned reviewers to be given the same weight as scores from unassigned reviewers, who may not have read an application in depth.

Dr. Hastie reviewed behavioral and psychological aspects of ranking and rating systems. He focused on the difficulties of distinguishing between top-rated applications within the current rating system and whether a ranking system would help resolve the issue. Concerns about the reliability and validity of ranking procedures were addressed. Dr. Hastie also noted the importance of designing studies to assess the outcome of peer review decisions as they are currently practiced.

Dr. Budescu's presentation involved statistical considerations to be taken into account when designing both scoring and ranking systems. He noted that many factors are considered desirable in any evaluation method—including reliability, validity, discrimination, transparency, and feasibility—and that sometimes improving one factor can adversely affect another factor. Dr. Budescu discussed the benefits of pair-wise comparisons when ranking a large number of applications and methods for aggregating individual scores or ranks produced during peer review meetings.

Dr. Saari gave the final presentation, discussing how changing voting methods can drastically change the order of preference for candidates in an election—a concept that could be applied to rank ordering of research grant applications. In particular, he examined what can happen when different procedures are used and how outcomes can be manipulated by changes in evaluation rules. Dr. Saari also briefly discussed a rank order pilot study being performed by the National Science Foundation that involves rank ordering of each application by seven reviewers, all of whom have read the applications in depth.

Conclusion

The speakers observed that studies on ranking methods were limited and they disagreed about the specific types of comparison that were likely to yield the best results. However, the speakers agreed that any direct ranking system would work best as an addition to rather than replacement of the current system. Furthermore, they noted the importance of evaluating the quality of peer review as it is currently practiced at NIH to ensure that the best science is indeed being selected for funding.

What People Think of NIH Peer Review Changes

Since our last issue, NIH posted the results of a second survey to gather feedback on recent NIH peer review enhancements. Reviewers and applicants as well as NIH review and program staff shared their thoughts.

Overall, applicants and reviewers are more satisfied with the new peer review system than the system in place before the Enhancing Peer Review initiative. Most respondents rated the peer review system as fair and consider themselves satisfied with the peer review process.

Some of the Specific Changes Assessed

- Shorter Applications
- Narrative Overall Impact Statement
- Nine-point Scoring Scale
- Bulleted Comments
- Single Resubmission Policy

Learn More

- Read the report online.

- Join the discussion of this report on the "Rock Talk" blog, which is hosted by the NIH Office of Extramural Research.

The Future

During the enhancement process, NIH made a commitment to continuously evaluate its peer review system to ensure that practices and policies uphold the core values of peer review. CSR has embraced this commitment and is working to further develop a science of peer review and a more rigorous means for assessing and guiding future changes to peer review and CSR practices and policies.

Subscribe to Peer Review Notes: www.csr.nih.gov/prnotes
Send comments or questions: PRN@csr.nih.gov

Center for Scientific Review
National Institutes of Health
U.S. Department of Health and Human Services